1956

An evaluation of critical thinking ability based on intelligence.

William J. Broderick
University of Massachusetts Amherst

AN EVALUATION OF CRITICAL THINKING ABILITY BASED ON INTELLIGENCE

BRODERICK - 1955

AN EVALUATION OF CRITICAL THINKING ABILITY BASED ON INTELLIGENCE

by

WILLIAM J. BRODERICK

A problem submitted in partial fulfillment
of the requirements for the Master of
Science Degree

University of Massachusetts

1956

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF GRAPHS

LIST OF GRAPHS

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

Volumes have been written about the great advances public education has made in the last one hundred and sixty years. We have advanced from the days when educational opportunities were available only to the more fortunate, to the present day, when classrooms are open to all. This change has provided us with a very literate citizenship. By means of magazines, newspapers, radios and televisions, the citizen of today is much more aware of the political, economic, social, and racial issues affecting his generation.

Present Day Educational Practices -- Our system of democracy has given the average person a chance to educate himself to a degree where he can be aware of the evils threatening his way of life. The question arises, does he have the ability to understand thoroughly the basic issues of these problems? It has been felt by many that despite our great strides in improving our educational facilities, we have not been able to make better citizens. For example, many feel that our public education teaches the pupils to read, but does not teach them to evaluate critically what they read. Similarly, pupils are aware of the political, economic, and social problems, but they lack the critical mind necessary to make intelligent, logical judgments. This inability leaves our citizens open to vigorous, persuasive orators and writers, who, in many cases, are seeking to undermine our democratic way of life.

We are living in an age in which speed is an important factor. The people of our country are constantly on the move. This condition makes a definite impression on our way of thinking. The voice of public opinion is a loud, but not always a correct, one. Not enough people take time to study controversial issues or to read deeply into the words of the press. Many quickly, and thoughtlessly, form their opinions and voice them with confidence backed only by self-assurance.

Problems of Our Educational System -- To step into this rapidly moving civilization and to check the pace of the human intellect is the problem confronting today's educators. A democracy needs and demands a citizenry versed in the problems it must face if it is to survive. However, this citizenry must be able, not only to comprehend all of the facts involved in a given issue, but also by a logical and critical method of reasoning to evaluate these facts before drawing a conclusion.

Definition of Critical Thinking -- Watson and Glaser maintain that critical thinking involves three factors:

"a) An attitude of wanting to have supporting evidence for opinions or conclusions before assuming them to be true.

b) Knowledge of the methods of logical inquiry which help determine the weight of different kinds of evidence and which help one to reach warranted conclusions.

c) Skill in employing the above attitude and knowledge."[1]

It is important to know where to get the facts necessary to form our opinions or to reach our conclusions. Our senses are the means or instruments whereby we obtain our knowledge of the external world. It would be folly to doubt our senses. However, it must be remembered that our senses can be, and often are, misleading. Another weakness here is that we are limited to the things we can see, touch, hear, smell, and taste, or to things which have happened or are happening to us. Our ability to relate facts learned by the previously mentioned methods to that which has occurred to cause them, or which may occur as a result of them, is limited. This limitation of our senses forces us, therefore, to accept facts on someone else's authority. Facts of history, geography and science are passed on and accepted as being true. It would be impossible for us, through our own experiences, to accumulate sufficient factual information to form true, accurate conclusions and opinions. Hence, it is necessary for us to read--it is necessary for us to think critically.

---

(1) Watson, Goodwin and Glaser, Edward Watson - Glaser Critical Thinking Appraisal Manual p. 8.

Qualities Necessary for Critical Thinking -- Ability to think logically is based on more than an assimilation of all the known facts. The person must have the intelligence to draw a logical conclusion from what he has accumulated Logic as an isolated subject is not found in the curriculum of today, but occasions for its use occur all through the day in every walk of life. Inside and outside of the classroom, students are required to make decisions or to draw conclusions by thinking logically. In many cases they seem to lack the ability to make a right judgment. This may be due to insufficient factual data, inadequate knowledge of the procedure to follow, and/or lack of intelligence to draw the correct conclusion from the assembled facts.

Reasons for Variations in Critical Thinking Ability -- In studying an individual's ability to think critically, there are several factors to be considered. The level of emotional maturity that a person has reached has an effect on his ability to think critically. The question arises whether or not a person has the stability to consider the facts involved and to judge from these facts and not from the outside forces that exert pressure to sway his decision. Even in the light of presented facts, many persons with strong prejudice will not alter their course of reasoning.

Another important phase in studying individuals is the thought pattern of a person. Though several people may seem to have the same degree of intelligence, some may have a much keener analytic mind. An analytic mind is characterized

by a pattern of thought that follows through from a given statement to a conclusion in a step by step manner with sound reasoning backing every step.

Of persons of equal intelligence, some show a process of logical thought in one field of study and very little in another. This usually occurs when they are vitally interested in a certain subject matter. Here their background of adequate knowledge helps them to draw their conclusions. Others have no specific strong field, but, because of a varied background in reading, develop critical habits of thought. Variations in critical thinking occur in groups of well-read people. These variations are due to differences in the reading range and comprehension of the individuals. A person with a wide range of reading and a high degree of comprehension usually has a sufficient background from which to draw his conclusions.

It can be said that some have a more critical mind by nature, while others develop the habits of critical thought in the same manner as they increase their knowledge. The habit of critical thinking can be improved and should be one of the goals of the modern school program.

This study is concerned with various types of critical thinking.

CHAPTER II

APPROACH TO THE PROBLEM

CHAPTER II

APPROACH TO THE PROBLEM

This problem is concerned with an attempt to discover if there is any appreciable difference in the quality of reasoning between pupils with an I.Q. range of 115-155 and those with a range of 60-100.

The Problem -- To evaluate the given test with the intent of locating if and where the difference in critical thinking occurs. The areas to be considered under critical thinking are inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments.

Subjects -- A heterogeneous group of 400 grade ten students. There has been no attempt to separate these students by age, sex, or course of study. Courses of study elected by this group include college preparatory, technical, general, commercial, drafting, home economics, welding, auto body, auto mechanics, cabinet making, and sheet metal.

Materials -- The Watson-Glaser Critical Thinking Appraisal, Form Am. Also the intelligence test scores of the 400 students taking the test. These I.Q. results are based on the Terman-McNemar Test of Mental Ability.

Procedure --

a) The 400 students who took the test were ranked by their scores on the intelligence test.

b) The top 25 percent are called Group A and the bottom 25 percent are Group B.

c) The number of correct answers for each question was tabulated for both Group A and Group B after the test had been given and scored.

d) The percent of pupils answering each question on the test correctly was determined and this data was summarized in a chart and the critical ratio determined to show the variance of the two groups.

e) The critical ratio is determined by:

1) finding the difference between the two groups

2) using Group A and the formula $\Sigma P_A = \sqrt{\frac{PQ}{N}}$

a) P is the percent of correct answers

b) Q is the difference between the correct answers or P and 100

c) N is the number of cases and in this study (100)

3) using Group B, repeat Step 2

4) following the formula $\Sigma d = \sqrt{(\Sigma P_A)^2 + (\Sigma P_B)^2}$ or more simply add the squares of the results of Step 2 and Step 3 and finding the square root of the result

5) finding the critical ratio is then $\frac{D}{\Sigma d}$ or the division of Step 1, the difference, by Σd, Step 4

This type of question by question analysis was used in order to locate the strength and weaknesses of the groups.

This information has been compiled and analyzed and is presented in Chapter III.

The Watson-Glaser Critical Thinking Appraisal has been used in this study as a measure of skill in the ability to think critically. Obviously, no conclusions which are based on the results of this test could be considered any more valid or reliable than the test itself. Therefore, before any results are appraised or conclusions drawn, it would be proper to evaluate the test as to its validity and reliability. Specifically, there are two questions which should be asked:

(1) Does the test actually measure critical thinking ability?

(2) Does the test measure the ability consistently?

Validity -- According to Wert, "Validity is defined as the degree to which a test measures what it is supposed to measure."[1] Applying this definition to the test used in this study, one should ask how accurately the test measures critical thinking ability, keeping in mind how this ability has been defined by the authors of this test.

In establishing the validity of the test used in this study, the authors of the test first selected various

---

(1) Wert, James Educational Statistics p. 205

exercises which they believed would measure the aspects of thinking which have been referred to and defined previously. These items were then submitted to a board of thirty-five "experts" selected on the basis of their extensive training in logic, language meaning, and scientific method. The board agreed unanimously that proficiency in individual aspects of thinking ability, which were involved in the test, would when considered collectively be indicative of critical thinking ability.

The authors further maintain that high correlation between success on individual items and on the total test yield further evidence in support of the validity of the key and the test itself.

Reliability -- When reference is made to the "reliability" of a test, the quality in question is the degree of consistency of the results which may be expected from repeated applications of the test.

The reliability of this test as a whole and of the separate sub-test scores has been determined by both the split-half and the inter-form method for several types of groups. When one test form is used, it has a reliability coefficient of .84.

CHAPTER III

DATA ON THE TWO GROUPS

CHAPTER III

DATA ON THE TWO GROUPS

It is the purpose of this chapter to present the data collected in the form of tables and graphs so that the comparisons can be more readily seen.

Intelligence Test Results -- The result of the intelligence testing was the only basis for selecting the two groups. The 400 students tested represented a normal tenth grade distribution with I.Q.'s ranging from 60 to 155 as can be seen in Graph 1. The mean I.Q. of these 400 students was 107.5 and the median was 108.1. There was a standard deviation of 15.2 for the group.

Group A and Group B formed two distinct ranges with no overlapping of I.Q. scores. These results are illustrated on Graph 2, where it can be seen that Group B had a range from 60 to 100, and Group A from 115 to 155.

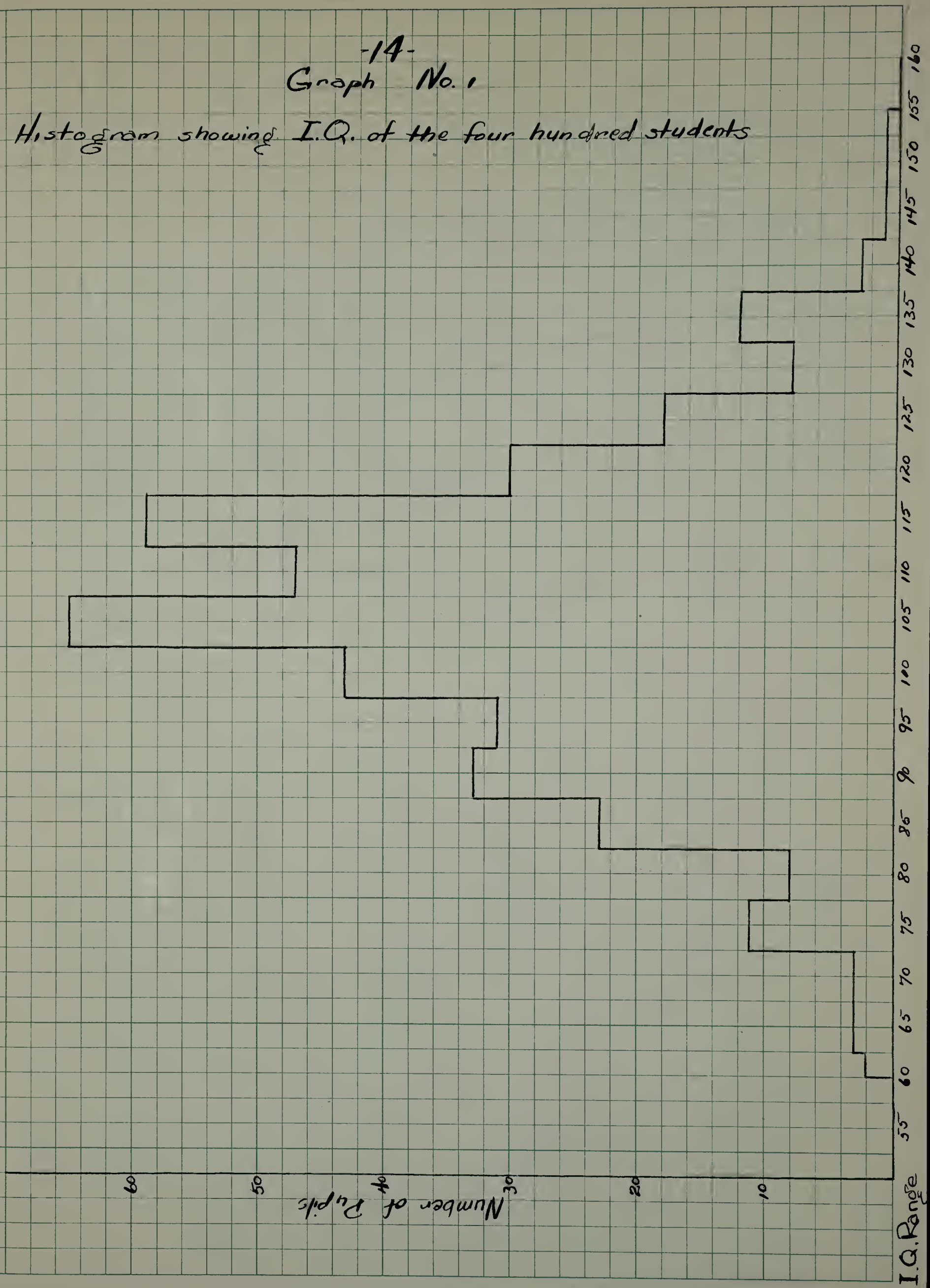
TABLE I

Comparing the I.Q.'s of Group A and Group B

	Group A	Group B
Nos. Pupils	100	100
Range I.Q.	115-155	60-100
Mean I.Q.	125.4	88.2
Median I.Q.	123.4	90.4
Standard Deviation	7.4	8.1

Graph No. 1

Histogram showing I.Q. of the four hundred students



Graph No. 2

A Comparison of the I.Q. Scores of Group A and Group B

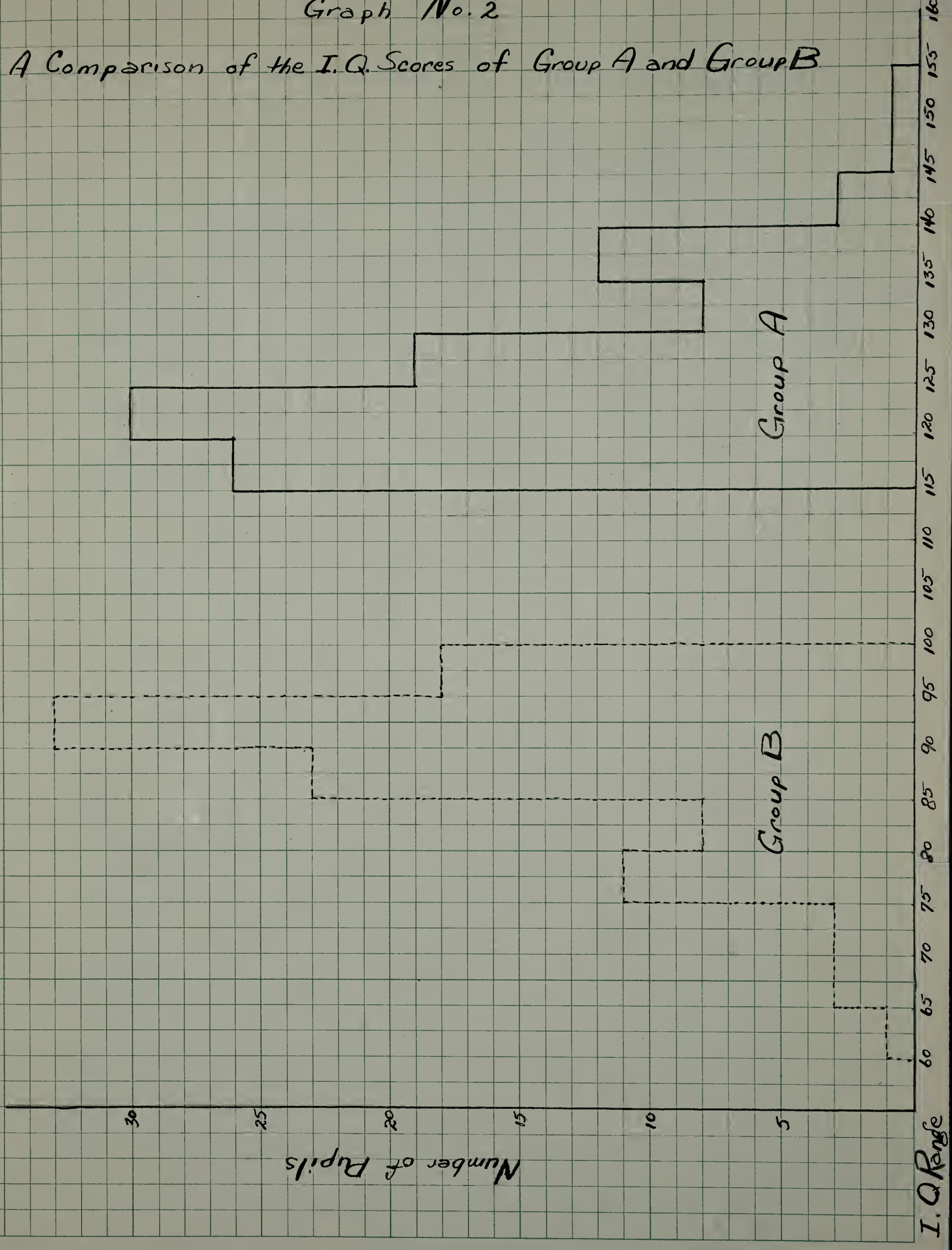


Table I shows the definite superiority of Group A in I.Q. There the mean and the median scores are more than 30 points greater, and the smaller standard deviation shows them to be a more concentrated group.

Evaluation of Raw Score Results -- The raw scores of the test have little or no interpretive significance. They must be evaluated against an established set of norms. Watson-Glaser in their manual for Critical Thinking Appraisal have provided a table listing the percentile norms for high school students. Under their classification "percentiles of 94 and above may be considered as "Level I: Very High," of 70-93 as "Level II: High," of 32-69 as "Level III: Average," of 8-31 as "Level IV: Low," and of 1-7 as "Level V: Very Low."[1]

Using this as a background against which to judge the raw score results, the 400 students tested had a percentile rank of 63, which puts them in Level III, an average group. Group A had a percentile rank of 87, which places them in Level II, High. Group B had a rank of 19, putting them in Level IV, Low.

Table II shows more clearly the comparative raw score results of the test for Group A and Group B. However, the most significant fact of this table is the percentile result,

---

(1) Watson, Goodwin and Glaser, Edward Watson-Glaser Critical Thinking Appraisal Manual p. 6.

which is based on the table provided in the Watson-Glaser manual. This shows the rank of these two groups compared with other high school students.

TABLE II

Group A and Group B Compared by Test Scores

	Group A	Group B
Raw Score Median	67.7	44.5
Raw Score Mean	67.3	45.1
Raw Score Standard Deviation	5.8	7.2
Percentile Rank	86	19

Summary -- The method of selection very evidently resulted in two distinct groups of intelligence with a minimum difference of fifteen intelligence quotient points between the highest student of the low group and the lowest student of the high group. This lack of overlapping is significant in the interpretation of the results in Chapter IV.

CHAPTER IV

RESULTS OF THE STUDY

CHAPTER IV

RESULTS OF THE STUDY

In this chapter the test will be broken into the five subtests and the groups will be compared item by item.

Test I - Inference -- The first subtest is Inference which the authors propose as "designed to sample ability to discriminate among degrees of truth or falsity or probability of certain inferences drawn from given facts or data."[1]

This subtest has twenty items and from Table III, the superiority of Group A can be seen.

TABLE III

TEST I	INFERENCE	20 ITEMS
	Group A	Group B
Raw Score Mean	11.9	5.9
Raw Score Median	11.8	5.7
Raw Score Standard Deviation	2.1	2.3

Table IV can be used to spotlight the strengths and weaknesses of the groups in this type of question-by-question analysis. Although statisticians vary on the size of the critical ratio that will denote significance, throughout this study any critical ratio over 3 will be considered significant.

(1) Watson, Goodwin and Glaser, Edward Watson-Glaser Critical Thinking Appraisal Manual p. 1.

On Test 1, which can be followed by referring to Table IV, there are four questions; 3, 7, 10, and 13 that do not give a significant critical ratio, i.e. the superior I.Q. students of Group A do not show a significant superiority on these questions. On the questions 3 and 13, both groups had very low percentage correct while on questions 7 and 10 the percentage correct follows more closely the results of the other items. Group A shows definite superiority on questions 9, 17, 18, and 19 where the critical ratio is 10, 13, 9, and 8 respectively.

Before seeking any reasons for these results a short explanation is necessary concerning the form used in answering these questions on Test 1.

In this test each exercise begins with a statement which is to be regarded as true. On the answer sheet the student has five choices to answer the question, they are: T, PT, ID, PF, and F.

T -- if inference is definitely true; that it follows from statement of fact given

PT -- if inference is probably true

ID -- insufficient data; you cannot tell from given facts whether the inference is likely to be true or false

PF -- if inference is probably false

F -- if inference is definitely false

TABLE IV

Item-by-Item Performance on Inference

Question	Percent Correct Group A	Percent Correct Group B	Group A Difference	Critical Ratio
1	93	50	+43	7.0
2	73	44	+29	4.0
3	29	32	- 3	0.4
4	47	23	+24	3.0
5	65	40	+25	4.1
6	40	13	+27	4.2
7	64	54	+10	1.2
8	61	37	+24	3.5
9	89	32	+57	10.3
10	51	38	+13	1.4
11	70	25	+45	7.1
12	72	24	+48	7.8
13	19	21	- 2	0.3
14	40	17	+23	3.0
15	46	13	+33	5.7
16	32	13	+19	3.6
17	90	22	+68	13.1
18	81	24	+57	9.2
19	66	16	+50	8.1
20	20	6	+14	3.0

The most notable feature of this test and apparently the cause of many mistakes for both groups is the answering of questions by basing reasoning on personal knowledge. The students did not refer to the given statement of fact to draw their conclusions but answered according to what their opinion might be on the topic.

Question 17, which gave Group A the greatest critical ration of the entire test, is "Tuberculosis can be cured." From everyday life this is an obviously true statement, but the given paragraph discusses only rate of tuberculosis between Whites and Negroes and makes no mention of the probability of cure.

Another notable fact, especially true with Group B, is that the pupils were more successful in answering the first question after each paragraph than they were in answering the succeeding ones. This feature is not as pronounced in Group A and might indicate that their higher I.Q. gives them a better retentive memory of the facts given in the paragraph.

Summary on Inference --

a) Many incorrect answers apparently caused by basing reason for answer on personal opinion instead of the facts given in the statement.

b() First question after the given paragraph has highest number correct with the last question having the lowest number correct.

c) In general the higher group was consistently better and maintained a definite superiority but in every case the lower group had some success.

Test 2 - Recognition of Assumptions -- Subtest 2 deals with the recognition of assumptions and contains 16 items. Its purpose is to test recognition of unstated assumptions in given assertions or propositions.

As will be seen in Table V, Group A maintains a marked superiority in raw score results although not as great a superiority as in the results of subtest 1.

TABLE V

TEST 2 RECOGNITION OF ASSUMPTIONS 16 ITEMS

	Group A	Group B
Raw Score Mean	12.5	8.32
Raw Score Median	12.9	8.2
Raw Score Standard Deviation	2.5	2.5

The test form used in this subtest has the answers under one of two columns. Either the assumption is "Made" or "Not Made." Now in this type of answer as in true-false questions, the problem of guessing has an affect on the results but to what degree is difficult to determine.

TABLE VI

Item-by-Item Performance on Recognition of Assumptions

Question	Percent Correct Group A	Percent Correct Group B	Group A Difference	Critical Ratio
21	99	76	+23	5.1
22	83	62	+21	3.2
23	70	45	+25	3.3
24	69	65	+ 4	0.6
25	68	24	+44	7.5
26	79	56	+23	3.2
27	72	41	+31	4.4
28	85	58	+27	4.2
29	69	49	+20	2.9
30	72	41	+31	4.6
31	63	48	+14	2.3
32	74	42	+32	4.5
33	74	29	+45	7.1
34	80	66	+14	2.3
35	81	40	+41	6.2
36	64	40	+24	3.0

As can be seen from Table VI, there is a critical ratio of less than 3 in 4 of the 16 items. More significant, however, is the fact that Group A does not show the marked superiority here that it did in Test 1. It has 7 as its greatest critical ratio in comparison to Test 1 where it went as nigh as 13.

The reason for the variance of the two groups may have been the diversity in the test form. In Test 2, the given statements are short, therefore, not requiring a retentive memory as in Test 1. As was the case in Test 1, it appears that the students answer the questions using their own opinions and not referring to the given statement before selecting the answer. The questions 24, 29, 31, and 34 in which there was a significant relationship between the groups show no other similarity or reason for causing this low critical ratio outside of the fact that they can be answered without reference to the given statement. In questions 25, 33, and 35 where the critical ratio was 7.5, 7.1, and 6.2 respectively, the reason for this difference is hard to explain.

Summary on Recognition of Assumptions --

a) Group A has marked superiority in raw score results.

b) The problem of students guessing at the answers is present in this test.

c) The critical ratio is not as great here as in Test 1.

d) Statement of the questions are much shorter in this test.

e) It appears that many answers are based on opinions instead of facts.

Test 3 - Deduction -- Subtest 3 is concerned with deductive type of reasoning. It is designed to sample ability to reason deductively from the given premises; to recognize the relationship of implication between propositions; to determine whether what seems as implication or necessary inference between one proposition and another is indeed such.

This is one of the longer subtests with 25 items. Table VII clearly shows that Group A continues to maintain a higher raw score.

As in Test 2, the answers are in two columns with the student to choose if the conclusion "follows" or "does not follow." This leaves again the difficulty of whether the student thought out or guessed at the answer.

TABLE VII

TEST 3 DEDUCTION 25 ITEMS

	Group A	Group B
Raw Score Mean	18.6	13.5
Raw Score Median	18.5	13.5
Raw Score Standard Deviation	2.3	2.6

TABLE VIII

Item-by-Item Performance on Deduction

Question	Percent Correct Group A	Percent Correct Group B	Group A Difference	Critical Ratio
37	97	72	+25	5.1
38	91	69	+22	4.2
39	21	23	- 2	0.3
40	97	87	+10	2.9
41	85	55	+30	4.2
42	25	26	- 1	0.1
43	86	55	+31	5.4
44	93	52	+41	7.3
45	100	70	+30	6.8
46	94	40	+54	9.8
47	90	61	+29	5.4
48	49	28	+21	3.2
49	28	31	- 3	0.4
50	97	81	+16	3.0
51	95	56	+39	7.4
52	97	68	+29	6.7
53	69	58	+11	1.6
54	94	84	+10	2.3
55	66	31	+35	5.4
56	60	46	+14	2.1
57	85	74	+11	1.9
58	30	23	+7	1.1
59	73	45	+28	4.2
60	58	40	+18	2.6
61	34	28	+ 6	0.9

This test on deduction shows a closer relationship between the two groups. Of the 25 items on the test, there are 11 that have a critical ratio below 3. This is by far the greatest percent shown in the first three tests. On the questions 44, 46, and 51, the critical ratio was 7.3, 9.8, and 7.4 respectively. It should also be noted that on question 45 Group A had a perfect score, the only question so answered on the entire test.

In searching for the reasons for the above results, it is again noticed that the given statements are short, putting less emphasis on reading comprehension. The important factor is the use of syllogisms in this test. This type of logical thinking is new to a tenth grade student. He is apparently unacquainted with the rules that govern this type of argumentation.

It is stated in the instructions for this subtest that the given two statements (premises) are to be considered as true. If the students did not keep this in mind they could be prejudiced in answering a certain question and thus draw the wrong conclusion.

Summary on Deduction --

a) Groups are closer on this test than on the previous two with Group A maintaining a higher raw score.

b) Problem of students guessing at the answers is also present in this test.

c)

c) Statement of the questions in this test are short putting less emphasis on reading comprehension.

d) Syllogisms are used in this test and could be the cause of many wrong answers if the students do not remember that the premises are to be taken as true.

Test 4 - Interpretation -- Subtest 4 is concerned with interpretation and is designed to sample ability to weigh evidence and to distinguish between unwarranted generalizations and probable inferrences which, though not conclusive or necessary, are warranted beyond reasonable doubt.

There are 24 items in this test and as can be seen in Table IX, the superiority of Group A on this test is not as large as in the previous tests.

TABLE IX

TEST 4 INTERPRETATION 24 ITEMS

	Group A	Group B
Raw Score Mean	16.3	12.6
Raw Score Median	16.3	12.4
Raw Score Standard Deviation	2.9	2.7

TABLE X

Item-by-Item Performance on Interpretation

Question	Percent Correct Group A	Percent Correct Group B	Group A Difference	Critical Ratio
62	85	73	+12	2.1
63	90	71	+19	3.5
64	65	40	+25	3.0
65	79	46	+33	5.0
66	58	48	+10	1.4
67	64	43	+21	3.0
68	37	30	+ 7	1.0
69	65	64	+ 1	0.1
70	70	37	+33	5.0
71	16	36	-20	3.2
72	86	73	+13	2.3
73	83	64	+19	3.0
74	40	35	+ 5	0.4
75	60	29	+31	4.0
76	96	80	+16	3.6
77	90	58	+32	5.5
78	50	38	+12	1.7
79	74	64	+10	1.5
80	47	43	+ 4	0.6
81	29	29	0	0.0
82	87	57	+30	5.0
83	85	51	+34	5.5
84	62	57	+ 7	0.7
85	57	43	+14	2.0

The form used in this test follows the pattern of the previous two tests with the answers in one of two columns. A given statement is assumed to be true and the person being tested is to judge if the proposed conclusion "follows beyond a reasonable doubt" or it "does not follow."

There are several notable facts to be drawn out of Table X. The first refers to question 71 where Group B has a favorable critical ratio of 3.2. This is the only question on the test where Group B has a significant superiority. Both groups had a low percentage score on this question, with Group A getting 16 and Group B 36. It is a difficult situation to analyze as the question 69, 70, and 71 are all drawn from the same sentence. Each question offers a possible solution to the problem created in the given sentence, but none of these questions could qualify as a definite solution to the problem because of the lack of information given in the sentence. It appears that both groups had in mind that one of the three given reasons must answer the question of the given sentence because both groups agreed that the conclusion of question 69 did not follow. If this assumption is true, it would follow that Group A selected question 71 and Group B, question 70 as the correct answer.

Group A shows no definite superiority on this test. Of the 24 items, 12 have a critical ratio of less than 3 and Group B has a favorable ratio of 3.2 on one question.

Group A has its greatest margin on questions 65, 70, 77, and 83 where its critical ratio is 5, 5, 5.5, and 5.5 respectively.

Summary of Interpretation --

a) Group A shows its least superiority on this test.

b) Group B had its only significant critical ratio on this test.

c) The statement of the questions in this test are also short.

Test 5 - Evaluation of Arguments -- The fifth and last subtest concerns the evaluation of arguments. It is designed to sample ability to distinguish between arguments which are strong and important to the question at issue and those which are weak and unimportant or irrelevant.

This test has 14 items and again the answers are in one of two columns. There is a series of questions and they are followed by three or four arguments. The arguments are to be regarded as true with the problem being whether it is a "Strong" or "Weak" argument.

TABLE XI

TEST 5 EVALUATION OF ARGUMENTS 14 ITEMS

	Group A	Group B
Raw Score Mean	11.1	7.2
Raw Score Median	11.1	7.2
Raw Score Standard Deviation	1.7	2.2

As can be seen from Table XI, Group A again has a considerable margin in the mean and median scores. There is also the greatest difference in standard deviation on this test with Group A having 1.67 which shows a concentrated score result.

TABLE XII

Item-by-Item Performance on Evaluation of Arguments

Question	Percent Correct Group A	Percent Correct Group B	Group A Difference	Critical Ratio
86	94	68	+26	5.0
87	95	64	+31	5.9
88	86	62	+24	4.0
89	64	43	+21	3.0
90	43	18	+25	4.0
91	52	41	+11	1.5
92	54	38	+16	2.3
93	57	32	+25	3.7
94	85	61	+24	4.0
95	94	46	+48	8.0
96	87	37	+50	8.0
97	85	59	+26	4.0
98	71	44	+27	4.0
99	94	59	+35	6.0

As can be seen on Table XII, there are only two questions 91 and 92 where there is a critical ratio of less than 3. On the other questions Group A shows a marked superiority especially on the questions 95 and 96 where they had a critical ratio of 8. A closer check of these four questions shows that they are of an entirely different type. The items 91 and 92 where the groups show a relationship are in answer to a question regarding the schedule of feeding babies. This is a question in which practical knowledge would favor over academic thus detracting from the strength of Group A. The items 95 and 96 where Group A showed its greatest margin are definitely academic questions and ones better answered by a student well versed in social studies. It pertains to the advantages and disadvantages of government owning and controlling industry.

Summary of Evaluation of Arguments --

a) Group A again showed a very wide margin in raw score results.

b) Group A has its smallest standard deviation on this test.

c) Statement of the questions in this test are also short.

d) Group A had its greatest margin on the academic questions.

CHAPTER V

SUMMARY AND CONCLUSIONS

CHAPTER V

SUMMARY AND CONCLUSIONS

Re-statement of the Problem -- This study is concerned mainly with finding out two things:

a) Is there a difference in the critical thinking ability of students with different intelligent quotients?

b) If there is a difference, in what areas of thinking does it appear?

In the study, two groups were chosen from a total of 400 pupils who had taken the Watson-Glaser Test. Group A had a range of intelligence from 115-155 and Group B had a range from 60-100.

Results and Conclusions --

a) The mean I.Q. of the 400 pupils taking the test was 107.5 and when divided into the top and bottom 25 percentile, Group had a mean I.Q. of 125.9 and Group B had 88.2.

b) On Test 3, Deduction, and Test 4, Interpretation, the two groups are closely related with Group A maintaining a small margin of superiority.

c) Group A shows a superiority in the tests on Inference, Recognition of Assumptions, and Evaluation of Arguments with its greatest margin on Inference.

d) In taking each test separately it is difficult to make a positive statement why or in what phase

of this subtest one group is strong or weak as the case may be with the limited number of items that are designed to test only for the type of reasoning concerning that subtest. Generalizations only can be made concerning materials within the test.

e) Throughout the entire test the most probable reason for incorrect answers was the prejudice of the students concerning certain questions. From the scores, one can assume that this fault is greater in the low group.

f) Many of the questions were such that when taken out of context the answers differed from the one to be used in the test. This could be another reason for many incorrect answers.

g) In Test 1, Inference, where Group A showed its greatest margin, the statement of the questions were long. This might indicate that the ability to read and comprehend the material is present to a higher degree in Group A.

h) In general, differences were differences of degree rather than quality. In every case Group B had some success but consistently below Group A.

i) This test is sufficient to show where the high I.Q. student is superior in regard to critical thinking. However, to determine the reasons for

the results, tests with a larger number of items would be necessary.

BIBLIOGRAPHY

Guilford, E. P. Fundamental Statistics in Psychology and Education. New York and London: McGraw-Hill Company, 1942.

Froehlich, C. P., Darley, J. G. Studying Students. Chicago: Science Research Associates, Inc. 1952 pp 216-259

Darley, J. G. Testing and Counseling in the High-School Guidance Program. Chicago: Science Research Associates, Inc. 1952. pp 78-87

Watson, G., Glaser, E. Watson-Glaser Critical Thinking Appraisal Manual. New York and Chicago: World Book Company, 1952

Leake, J. D., Durost, W. N. "Curriculum and Instructional Implications of Test Results" Test Service Bulletin No. 75 World Book Company

Shanner, W. A Guide to Logical Thinking. A Life Adjustment Booklet. Science Research Associates, Inc. 1954.

Wert, James Educational Statistics. New York: McGraw-Hill, 1938. p 205.

Approved by:

Albert W. Pursus

Edward H. Fenner

(Problem Committee)

Submitted: June, 1956

WATSON-GLASER CRITICAL THINKING APPRAISAL *Form* AM

by GOODWIN WATSON

Professor of Education, Teachers College, Columbia University

and EDWARD MAYNARD GLASER

Consulting Psychologist, Rohrer, Hibler, and Replogle, Los Angeles

DIRECTIONS:

This booklet contains several different types of tests designed to find out how well you are able to reason analytically and logically.

Do not turn this page until instructed to do so. Do not make any marks on this test booklet. All answers are to be marked on the separate Answer Sheet provided. If you wish to change an answer, be sure to erase your old answer completely.

Published by World Book Company, Yonkers-on-Hudson, New York, and Chicago, Illinois. Printed in U.S.A. WGCTA : AM–2

TEST 1. *Inference*

DIRECTIONS. An *inference* is a conclusion which a person draws from certain observed or supposed facts. Thus, from the electric light visible behind the window shades and from the sound of piano music in a house, a person might *infer* that someone is at home. But this inference may or may not be correct. Possibly the people in the house went out leaving the lights on, and the piano music could be coming from a radio or phonograph they left playing.

In this test each exercise begins with a *statement of facts* which you are to regard as true. After each statement of facts you will find several possible inferences — that is, inferences which some persons might make from the stated facts. Examine each inference separately, and make a decision as to its degree of truth or falsity.

On the Answer Sheet you will find for each inference spaces marked with the letters T, PT, ID, PF, and F. For each inference make a mark on the Answer Sheet under the appropriate letter as follows:

T — if you think the inference is *definitely true;* that it properly follows from the statement of facts given.

PT — if, in the light of the facts given, you think the inference is *probably true;* that there is better than an even chance that it is true.

ID — if you decide that there are *insufficient data;* that you cannot tell from the facts given whether the inference is likely to be true or false.

PF — if, in the light of the facts given, you think the inference is *probably false;* that there is better than an even chance that it is false.

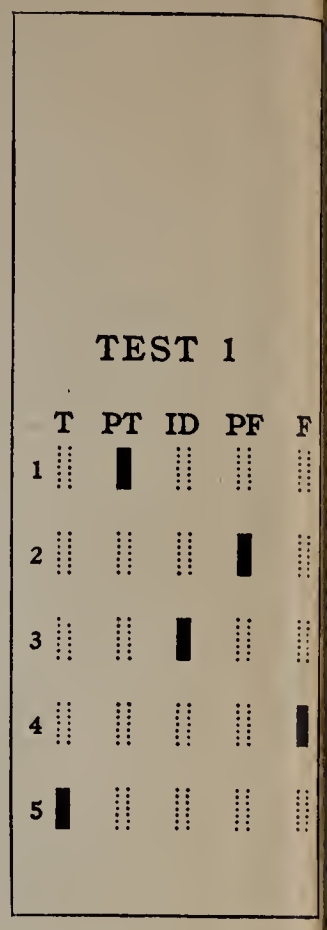
F — if you think the inference is *definitely false;* that it cannot possibly be drawn from the facts given or that in some manner it contradicts the facts.

Sometimes, in deciding whether an inference is probably true or probably false, you will have to use certain commonly accepted knowledge or information which practically every person knows. This will be illustrated in the example which follows.

Here is the example; the correct answers are indicated in the block at the right.

EXAMPLE. A thousand eighth-grade students recently attended a voluntary week-end conference in a Midwestern city. At this conference questions of race relations and means of achieving lasting world peace were discussed, since these were the problems the students felt to be most vital today.

1. As a group, the students who attended this conference had a keener interest in humanitarian or broad social problems than most eighth-grade students have.....
2. The majority of these students were between the ages of 17 and 18.....
3. The students came from all sections of the country.....
4. The students came to discuss trade-union problems.....
5. Some eighth-grade students felt that discussion of race relations and means of achieving world peace might be worthwhile.....



In the above example, inference 1 is *probably true* (PT) because (as is common knowledge) most eighth-grade students are not likely to evidence such serious concern with broad social problems.

Inference 2 is *probably false* (PF) because (common knowledge) there are relatively few eighth-grade students in the United States between 17 and 18 years of age.

There is no evidence for inference 3. Thus there are *insufficient data* (ID) for making a judgment in the matter.

Inference 4 is *definitely false* (F) because it is given in the statement of facts that race relations and means for achieving world peace were the problems discussed.

Inference 5 necessarily follows from the given facts; it therefore is *true* (T).

In the exercises which follow, more than one of the inferences from a given statement of facts may be true (T), or false (F), or probably true (PT), or probably false (PF), or have insufficient data (ID) to warrant any conclusion. That is, you are to consider each inference by itself.

Make a heavy black mark in the space under the letter that you think best describes each inference. If you change an answer, erase thoroughly. Make no extra marks on the answer sheet.

Go on to the next page.

An English teacher arranged for the students in one of her classes to see the movie *Great Expectations*, while the students in other classes studied the book itself, without seeing the picture. Tests to measure appreciation and understanding of the story were administered immediately upon completion of each type of instruction. On all tests the class which was taught with the aid of the movie did better. The class which saw the movie became so interested that before the semester was over most of those students read the book, entirely on their own initiative.

1. The tests to measure appreciation and understanding of the story were administered both to the students who saw the picture and to those who only studied the book.....

2. The children who were taught with the aid of the motion picture were required to read the book before the end of the semester.....

3. Pupils who see movies instead of reading books lose interest in reading.....

4. Most of the children in the class which saw the picture would have preferred to study the book *Great Expectations* in the usual way without the aid of the movie.....

5. The teacher who conducted the experiment will hereafter try to use motion pictures when they are available, as an aid in teaching literary appreciation.

6. Pupils can learn more about any given subject from motion pictures than they can from books.....

The first newspaper in America, edited by Ben Harris, appeared in Boston September 25, 1690, and was banned the same day by Governor Simon Bradstreet. The editor's long fight to continue his little paper and print what he wished marks an important episode in the continuing struggle to maintain a free press.

7. The editor of the first American newspaper died within a few days after his paper was banned.....

8. Governor Bradstreet felt he had the legal authority to ban Ben Harris's paper.....

9. The editor of this paper wrote articles against taxes of the kind which later brought about the "Boston Tea Party."

10. Ben Harris was a man of persistence in holding to some of his interests and convictions.....

Some time ago a crowd gathered in Middletown, Mississippi, to hear the new president of the local Chamber of Commerce speak. He said, "I am not asking, but demanding, that labor unions accept their full share of responsibility for civic betterment and community interests. I am not asking, but demanding, that they join the Chamber of Commerce." The listening representatives of the Central Labor Unions applauded enthusiastically. Three months later all the labor unions in Middletown were represented in the Chamber of Commerce, where they served enthusiastically on committees, spoke their minds, and participated actively in the civic betterment projects.

11. Both the labor union representatives and the other members of the Chamber of Commerce came to recognize one another's problems and viewpoints better through their Chamber of Commerce contacts.

12. Labor unions' participation in the Middletown Chamber of Commerce has largely eliminated worker-management disputes in that town.....

13. The active participation of the labor unions caused friction at the meetings of the Chamber of Commerce.

14. The union representatives soon regretted having accepted the invitation to participate in the Chamber of Commerce.....

15. Many of the Chamber of Commerce members came to feel that their president had been unwise in asking the union representatives to join the Chamber.....

16. The representatives of the Central Labor Unions joined the Chamber of Commerce against the desires of the great majority of their membership.....

Studies have shown that there is relatively much more tuberculosis among Negroes in the United States than among whites. There is no difference, however, in rate of tuberculosis between Negroes and whites who have the same level of income. The average income of whites in the United States is considerably higher than the average income of Negroes.

17. Tuberculosis can be cured.....

18. Raising the economic level of Negroes would reduce tuberculosis.....

19. Tuberculosis is less prevalent among Negroes with relatively high incomes than among Negroes with relatively low incomes.....

20. Whether a white person is rich or poor makes no difference in the likelihood of his getting tuberculosis.

Go on to the next page.

TEST 2. *Recognition of Assumptions*

DIRECTIONS. An *assumption* is something supposed or taken for granted. When someone states, "I'll graduate in June," he takes for granted or assumes that he will be alive in June, that he will remain in school until that time, that he will pass his courses, and similar things.

Below are a number of statements. Each statement is followed by several proposed assumptions. *You are to decide for each assumption whether it necessarily is taken for granted in the statement.*

If you think the given assumption is taken for granted in the statement, make a heavy mark between the dotted lines under "ASSUMPTION MADE" in the proper place on the Answer Sheet. If you think the assumption is not necessarily taken for granted in the statement, make a heavy line under "ASSUMPTION NOT MADE" on the Answer Sheet.

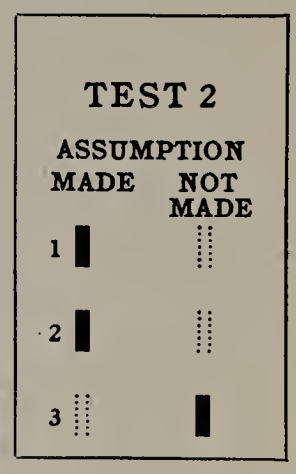
Below is an example: the block at the right shows how these items should be marked on the Answer Sheet.

If you do not see why the answers marked are right, ask the examiner to explain. In some cases more than one of the given assumptions is necessarily made; in other cases none of the given assumptions is made.

EXAMPLE. STATEMENT: "We need to save time in getting there, so we'd better go by plane."

PROPOSED ASSUMPTIONS:

1. Going by plane will take less time than going by some other means of transportation. (It is assumed in the statement that greater speed of a plane over other means of transportation will enable the group to get to their destination in less time.) .....
2. It is possible to make plane connections to our destination. (This is necessarily assumed in the statement, since, in order to save time by plane, it must be possible to go by plane.) ....
3. Travel by plane is more convenient than travel by train. (This assumption is not made in the statement — the statement has to do with saving time, and says nothing about convenience or about any other specific mode of travel.) .....



STATEMENT: "Let us immediately build superior armed force and thus keep peace and prosperity."

PROPOSED ASSUMPTIONS:

21. If we have superior armed force, that will insure the maintenance of peace and prosperity.....
22. Unless we increase our armaments immediately we shall have war.....
23. We now have peace and prosperity.....

STATEMENT: "A wise man will save at least twelve dollars each week out of his earnings."

PROPOSED ASSUMPTIONS:

24. No fools have sense enough to save twelve dollars a week.....
25. A person needs to be wise in order to save twelve dollars a week.....

STATEMENT: "Even if all the wealth in the country suddenly were to be distributed equally, some people soon would again become rich and others poor."

PROPOSED ASSUMPTIONS:

26. The real causes of wealth and poverty would not be much affected by such Socialism.....
27. Our present economic system is better than such Socialism.

STATEMENT: "Mary isn't going to invite John to her party."

PROPOSED ASSUMPTIONS:

28. Mary hasn't yet had her party.....
29. Mary now doesn't like John.....
30. The party will be at Mary's house.....

STATEMENT: "Live in the city of Zenith — lowest taxes."

PROPOSED ASSUMPTIONS:

31. Efficient management of a city implies lower taxes.
32. An important consideration in deciding where to live is avoidance of high taxes.....
33. The people of Zenith are content with their present city government.....

STATEMENT: "Our school is fortunate in having all American pupils, so we have no race problems."

PROPOSED ASSUMPTIONS:

34. American pupils do not present any race problems.
35. If we practiced democracy, there would be no race problem.....
36. A school is unfortunate if its pupils are of varied nationalities.....

Go on to the next page.

TEST 3. *Deduction*

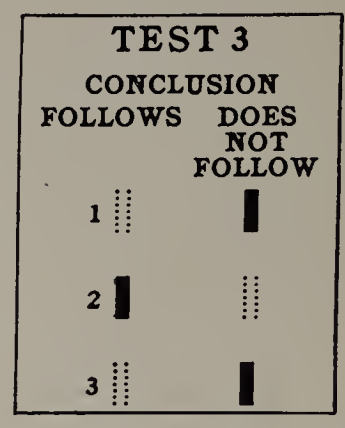
DIRECTIONS. Each exercise below consists of two statements (premises) followed by several proposed conclusions. For the purposes of this test, consider the two statements in each exercise as true without exception. Read the first conclusion beneath the statements, and if you think it *necessarily follows from the statements given,* answer by making a heavy black mark between the pair of dotted lines under "CONCLUSION FOLLOWS" in the corresponding blank on the Answer Sheet. If you think it is *not* a *necessary* conclusion from the given statements, then put a heavy black mark under "CONCLUSION DOES NOT FOLLOW," even though you may believe it to be true from your general knowledge.

Likewise read and judge each of the other conclusions. Try not to let your prejudices influence your judgment — just stick to the given statements and judge each conclusion as to whether it necessarily follows from them. Mark all your answers on the Answer Sheet.

Here is an example; the block at the right shows how your answers should be marked on the Answer Sheet.

EXAMPLE. Some holidays are rainy. All rainy days are boring. Therefore —

1. No clear days are boring. (The conclusion does not follow, as you cannot tell from these statements whether or not clear days are boring and some may be.) .....
2. Some holidays are boring. (The conclusion necessarily follows from the statements, since, according to them, the rainy holidays must be boring.) .....
3. Some holidays are not boring. (The conclusion does not follow from the statements even though you may know that some holidays are very pleasant.).....



All musicians are temperamental. Some musicians are not proud. Therefore —

37. All temperamental people are musicians.....
38. No proud people are temperamental.....
39. Some proud people are musicians.....

No jockey is a heavyweight boxer. All heavyweight boxers are large men. Therefore —

40. No jockey is a small man.....
41. No heavyweight boxer is a small man.....
42. Jockeys are small men.....

Some cannibals are sincere idealists. All cannibals are fanatics. Therefore —

43. Some sincere idealists are fanatics.....
44. Some fanatics are sincere idealists.....
45. No fanatics are sincere idealists.....
46. All fanatics are cannibals.....

All mice that are injected with substance "A" develop disease "X." Mouse #24 was not injected with substance "A." Therefore —

47. Mouse #24 did develop disease "X."....
48. Not all mice with numbers between 20 and 30 were injected with substance "A."....
49. Mouse #24 did not develop disease "X."....

No Republican is a Democrat. All Democrats favor prosperity. Therefore —

50. Republicans favor prosperity.....
51. No Republican opposes prosperity.....
52. No Democrat opposes prosperity.....
53. No Republican favors prosperity.....

All Jews feel friendly toward the State of Israel. David feels friendly toward the State of Israel. Therefore —

54. David is not friendly toward the Arabs.....
55. David is Jewish.....
56. Some non-Jews also feel friendly toward the State of Israel.....

If an adult has the ability to give love to others, he must have received love as a child. Some adults did not receive love when they were children. Therefore —

57. Some adults do not have the ability to give love to others.....
58. If an adult received love as a child, he has the ability to give love to others.....

If a person is superstitious, he believes fortunetellers. Some people do not believe fortunetellers. Therefore —

59. No superstitious person doubts fortunetellers.....
60. If a person is not superstitious, he will not believe fortunetellers.....
61. If a person believes fortunetellers, he is superstitious.....

Go on to the next page.

TEST 4. *Interpretation*

DIRECTIONS. Each exercise below consists of a short paragraph followed by several proposed conclusions.

For the purpose of this test assume that everything in the short paragraph is true. The problem is to judge whether or not each of the proposed conclusions logically follows beyond a reasonable doubt from the information given in the paragraph.

If you think that the proposed conclusion follows *beyond a reasonable doubt* (even though it may not follow absolutely and necessarily), then make a heavy black mark between the appropriate dotted lines under the "CONCLUSION FOLLOWS" column on the Answer Sheet. If you think that the conclusion does *not* follow beyond a reasonable doubt from the facts given, then make a mark under "CONCLUSION DOES NOT FOLLOW."

In some cases more than one of the proposed conclusions may follow; in other cases none of the conclusions may follow.

A report of the U. S. Census states that during 1940 there were approximately 1,656,000 marriages and 264,000 divorces granted in the United States.

62. Getting a divorce is a quick and easy matter in the United States.....

63. If the above ratio still holds true, then about six times as many people get married each year as get divorced.....

64. The divorce rate in the United States is much too high.....

Victims of radiation sickness (for example, after an atomic explosion) are likely to die of anemia because the blood-building properties of the bone marrow are damaged. In everyday medical practice, X-ray dosages have to be worked out with utmost care to keep the patient from falling prey to radiation sickness. Experimenting on rabbits, Dr. Leon Jacobson found that when the spleen and appendix were protected with lead, the animals survived what would otherwise have been a fatal overdose of X rays. The undamaged spleen and appendix make enough blood to enable the damaged tissue to recover.

65. If from the blood-forming organs a substance could be isolated which would speed an individual's recovery from radiation sickness, that substance probably would also enable X-ray patients to take heavier doses.....

66. Dr. Jacobson's experiments on rabbits should be tried on a sufficiently large scale with people to see whether the same results would hold true.....

Usually I fall asleep promptly, but about twice a month I drink coffee in the evening; and whenever I do, I lie awake and toss for hours after I go to bed.

67. My problem is mostly mental; I am over-aware of the coffee when I drink it at night, anticipating that it will keep me awake, and therefore it does.....

68. I don't fall asleep promptly after drinking coffee at night because the caffeine in coffee stimulates my nervous system for several hours after drinking it...

At the end of the semester the pupils in Mr. Black's class averaged 10 points higher than the pupils in Miss Walter's class on the same geometry test. Mr. Black and Miss Walter used a somewhat different method of teaching geometry.

69. Mr. Black probably is a better teacher than Miss Walter.....

70. The pupils in Mr. Black's class were brighter as a group than the pupils in Miss Walter's class, and therefore they learned more easily.....

71. The method used by Mr. Black in teaching geometry was superior to the method used by Miss Walter...

When Great Britain began to offer free public medical service, the government was surprised because far more people than they had expected came for eyeglasses and dental work.

72. People who previously had neglected their eyes and teeth now chose to have such treatment.....

73. People who didn't really need these services sought them because they were free.....

74. People in Great Britain previously had been careless about the state of their eyes and teeth.....

75. The British public was pleased with the government health program.....

Go on to the next page.

TEST 4. *Interpretation* (*Continued*)

The *Los Angeles Times* made a survey of the number of men and women drivers involved in automobile accidents in the Los Angeles area during a given period of time. They found that men drivers were involved in 1210 accidents while women drivers were involved in only 920 accidents.

76. If the survey figures constitute a representative sample, men drivers are involved in accidents more frequently than women drivers in the Los Angeles area. . . . . .
77. More men than women drive cars in the Los Angeles area. . . . . .
78. Women are safer drivers than men in the Los Angeles area. . . . . .

Intelligence tests show that Negro children in Northern cities surpass Negro children in Southern cities but do not score as high as white children in Northern cities.

79. White children as a group score higher because they are born with higher native intelligence than Negro children. . . . . .
80. The Negro families who moved to the North are on the average more intelligent than those who remained in the South. . . . . .
81. Northern Negroes receive better schooling than Southern Negroes, which in turn influences performance on the tests. . . . . .

The history of the last two thousand years shows that wars have become steadily more frequent and more destructive, the twentieth century being the bloodiest on record.

82. Mankind has not advanced as much in the art of keeping peace as it has in the science of waging war.
83. Wars are caused by basic traits of selfishness, greed, and pugnacity, which are rooted in human nature. .
84. Increased industrialization, competitiveness, and improved weapons bring on increasingly frequent wars. . . . . .
85. There will be increasingly frequent future wars, and they will become steadily more destructive than past wars. . . . . .

Go on to the next test.

TEST 5. *Evaluation of Arguments*

DIRECTIONS. In making decisions about important questions it is desirable to be able to distinguish between arguments that are *strong* and those which are *weak* in so far as the question at issue is concerned.

Strong arguments must be both important and directly related to the question.

Weak arguments may not be directly related to the question, even though they may be of great general importance; or they may be of minor importance; or they may be related to trivial aspects of the question.

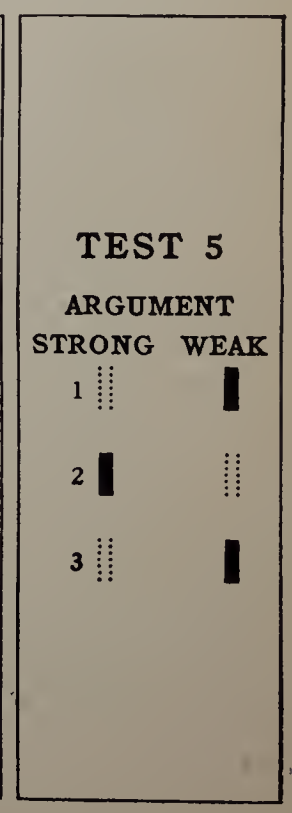
Below is a series of questions. Each question is followed by three or four arguments. *For the purpose of this test you are to regard each argument as true.* The problem then is to decide whether it is a STRONG argument or a WEAK argument.

You are to answer by making a heavy mark on the Answer Sheet under "STRONG" if you think the argument is strong, or by making a heavy mark under "WEAK" on the Answer Sheet if you think the argument is weak. When evaluating an argument, judge it on its own merit; try not to let counter-arguments or your own attitude toward the question influence your judgment. Judge each argument separately. In some questions all the arguments may be STRONG, in others all may be WEAK.

Here is an example. The block at the right shows how these arguments should be marked on the Answer Sheet. Study them carefully until you know just what is expected of you. Note that the argument is evaluated as to how well it supports the side of the question indicated.

EXAMPLE. Should all young men go to college?

1. Yes; college provides an opportunity for them to learn school songs and cheers. (This would be a silly reason for spending years of one's life in college.)
2. No; most young men profit more from work experience than from college classes. (According to the directions, we must accept this argument as true; hence it is a strong and important one against all young men going to college.) . . . . . .
3. No; excessive studying permanently warps an individual's personality. (This argument, although of great general importance when accepted as true, is not directly related to the question, because attendance at college does not necessarily require excessive studying.) . . . . .



Go on to the next page.

Remember that for the purpose of this test each argument is to be regarded as true.

Can rich and poor people who happen to oppose each other at law obtain approximately equal justice from the courts?

86. No; a rich person can hire better lawyers and technical experts, pay for the time of more witnesses, and continue the fight in higher courts.....

87. No; rich people win the majority of their lawsuits against poor people.....

Should married women be eligible for employment as public school teachers if they are otherwise qualified?

88. No; there are more single women in our country than there are school-teaching jobs.....

89. Yes; women tend to become better teachers after marriage.....

90. No; a mother's first responsibility is to her own children.....

Should infants be fed by regular schedule rather than whenever they seem to be hungry?

91. No; babies know best when they are hungry and ready to eat.....

92. Yes; children must sooner or later learn that they can't always have their own way.....

93. Yes; a regular schedule is easier for the parents....

Should the government take over all the main industries in the country, employ all who want to work, and offer the products at cost prices?

94. No; so much concentration of economic and bureaucratic power in government would undermine our personal and political freedom.

95. No; elimination of competition and the profit motive would result in much less initiative for production of useful new goods and services.....

96. Yes; the government already operates post offices, highways, parks, military forces, public health services, and other public services.....

Should groups in this country who are opposed to some of our government's policies be allowed unrestricted freedom of press and speech?

97. Yes; a democratic state thrives on free and unrestricted discussion, including criticism.....

98. No; if given full freedom, opposition groups would disunite the American people, weaken our position, and ultimately lead to loss of our democracy.....

99. No; the countries opposed to our form of government do not permit the free expression of our point of view in their territory.....

Go back and check your work.